A Appendix

A.1 Reasoning for image size

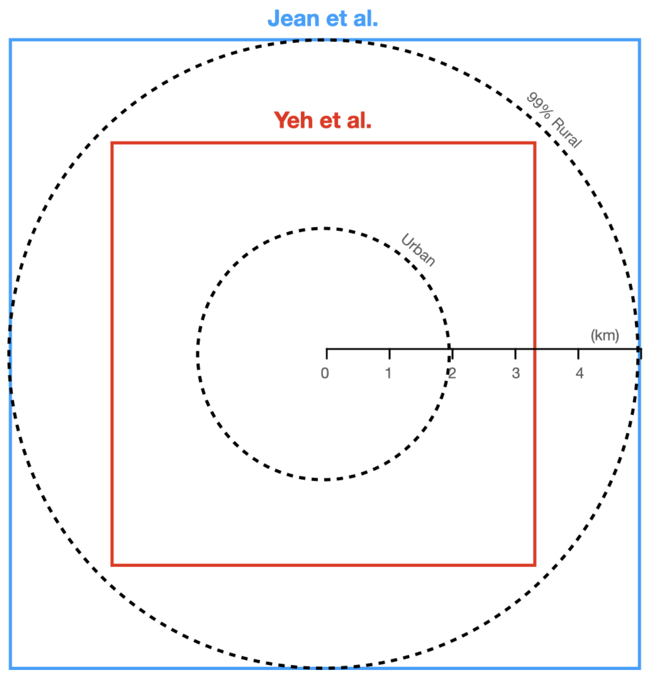


Figure 6: Input image dimensions used in earlier works compared to maximum displacement for urban and rural DHS coordinates. The limit for the 1% of rural points which are displaced up to 10 km is omitted. For this article, we elected to follow the example set by Yeh *et al.* (2020).

Selecting an appropriate input size for our model is a trade-off between including information about the relevant neighborhood and excluding diluting information about surrounding areas. Due to the random displacement mentioned in Section 2.1, 42 % of the rural cluster centers end up completely outside of our selected image frames (see Figure 6). Note that the centroid of a survey cluster being completely outside of the image is not the same as no part of the neighborhood being inside the image. At most, the centroid will be 1.64 km outside the image border. The exception is the one percent of rural locations which have been displaced by up to 10 km. However, all the urban centroids will be fully within the image frame and at least 1.36 km inside the image border. On one hand, you obviously want to include the cluster center in the image, but on the other, you don't want to include too much of the surrounding area. All images will contain locations at least 4.75 km away from the cluster center and due to the nature of poverty, such a distance might already cover vastly different levels of wealth, especially in urban areas. A bigger coverage area will exacerbate this problem. As the ResNet models we use have an input size of 224×224 pixels and each pixel covers 30×30 meters, 6.72×6.72 km is a size that fits without any modifications. Due to this fact and that it has some precedence in the work done by Yeh *et al.* (2020), which we have largely based our models on, we selected it.

A.2 Window configurations

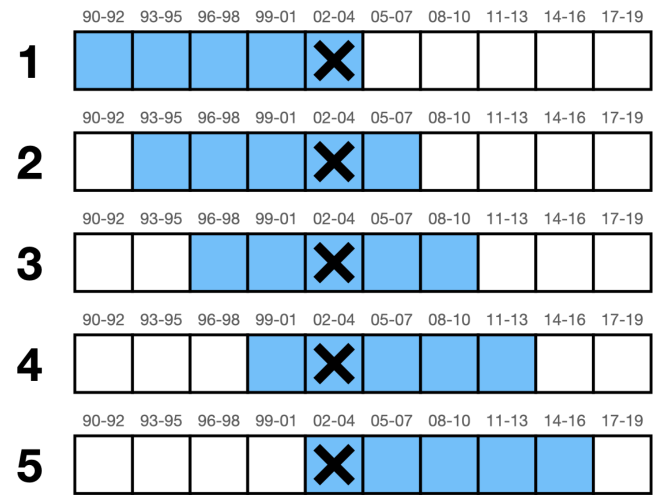


Figure 7: The five possible window configurations for a FIVE-FRAME model when predicting the wealth of a neighborhood in 2003. The blue squares represents the input-window sequence given to the model.

Two different versions of the MULTI-FRAME architecture was implemented: a FIVE-FRAME model, which considers a sequence of five image frames, and a TEN-FRAME, which considers ten frames. As each frame represent a three-year composite this means that the FIVE-FRAME model has access to 15 years worth of imagery, whereas the TEN-FRAME model considers the full 30 years time span between 1990 to 2019 for which we have data. A major reason for using a FIVE-FRAME model in addition to the TEN-FRAME model is that a model covering the entire time span for which we have data will always have the same output indices corresponding to the same time-frames. The first output of the TEN-FRAME model will always represent the years 1990-1992, the second will be 1993-1995 and so on. As only a subset of countries are surveyed within each time frame it is therefore possible that such a model might overfit the predictions for a given output index based on the surveyed countries in this time frame. As an example, for the first time-frame (90-92) over half of the surveyed clusters were located in Egypt, a relatively rich country, which might lead a model to overestimate the wealth of other locations in that time-frame. For the model which covers only 15 years out of 30 we can somewhat circumvent this problem by considering different time spans for input/output, as illustrated in Figure 7. For the five frame model, the first output could represent 90-92, but it could also be 93-95 or 02-04 depending on which time span we are considering. By showcasing that the FIVE-FRAME model also significantly outperforms the SINGLE-FRAME model, we wish to show that this concept do not explain the full advantage of using a MULTI-FRAME architecture.

A.3 Splitting clusters in OOA

For the Out-of-area experiment, the idea was to randomly assign each cluster to a cross-validation fold with equal probability. However, due to many survey points being located so close to each other, this would result in the same location appearing in multiple images. If a model is trained on one of these images and evaluated on the other then this might constitute ”peaking” at the test set. To avoid this, clusters that were located close enough to one another for an overlap to occur were grouped into ”collections” using DBSCAN. The list of collections was then sorted in descending order by the number of clusters they contained before each collection was greedily assigned to the least populated fold.

Originally, these collections would become quite large due to the density of collected data points forming long chains of overlapping images. As an example, almost all of the surveyed clusters in Egypt, which are densely situated along the narrow Nile river, ended up in the same collection. To counteract this, a small set of clusters were manually removed from the data set (0.7%) to ”break the chains” of these larger collections before applying the grouping algorithm. This resulted in a more even distribution, whereas folds would otherwise be dominated by a single continuous region.

A.4 Aggregated performance

A model's performance can be further boosted by aggregating predictions to the district level. When comparing the mean predictions versus the mean surveyed IWI for each district, our TEN-FRAME OOC model is are able to explain 78% of pooled variation. This boost to performance can likely be explained by prediction and surveying errors canceling out via averaging. This relationship can be further seen in Figures 9-12.

A.5 Comparing to HDI

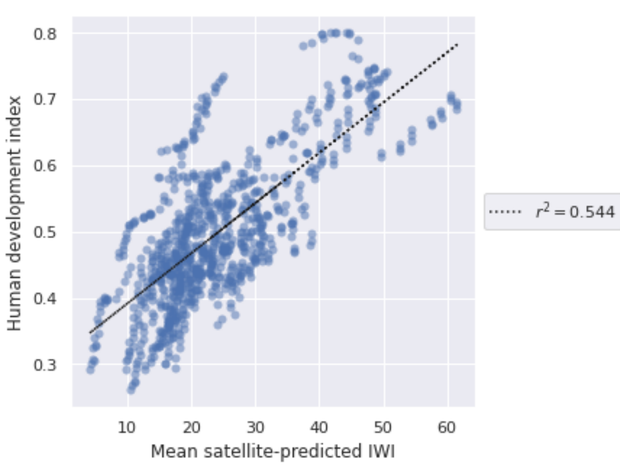
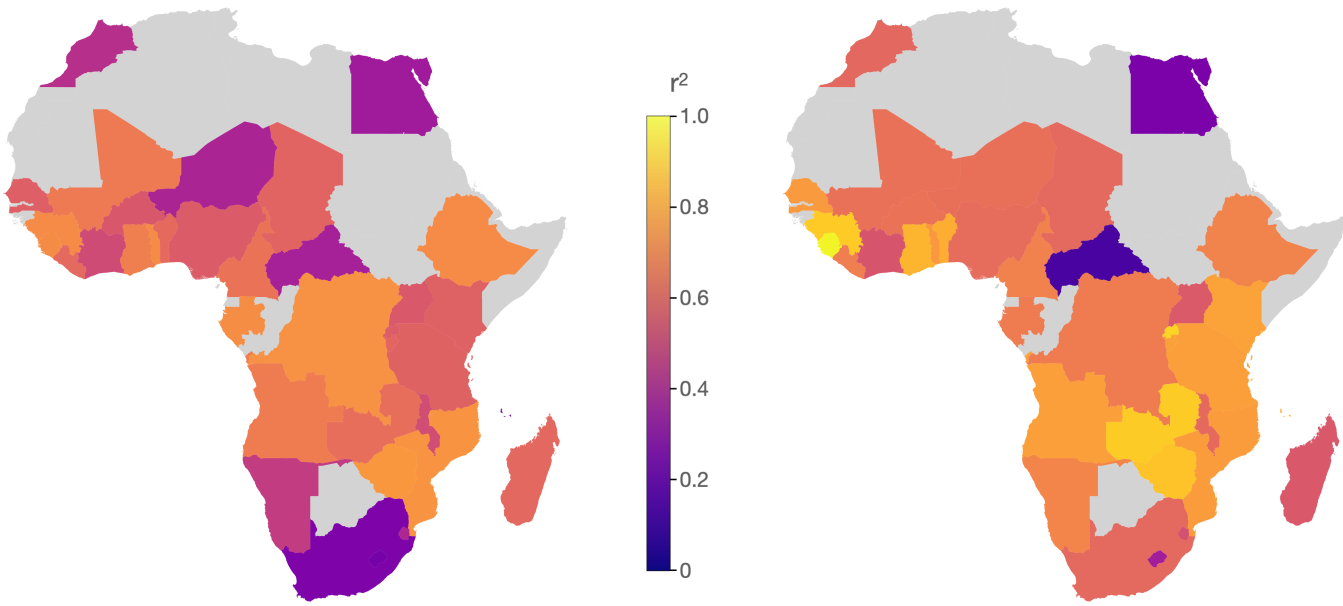


Figure 8: The mean satellite-predicted IWI for a corresponds well with the Human development index (HDI) for the corresponding country-year pair.

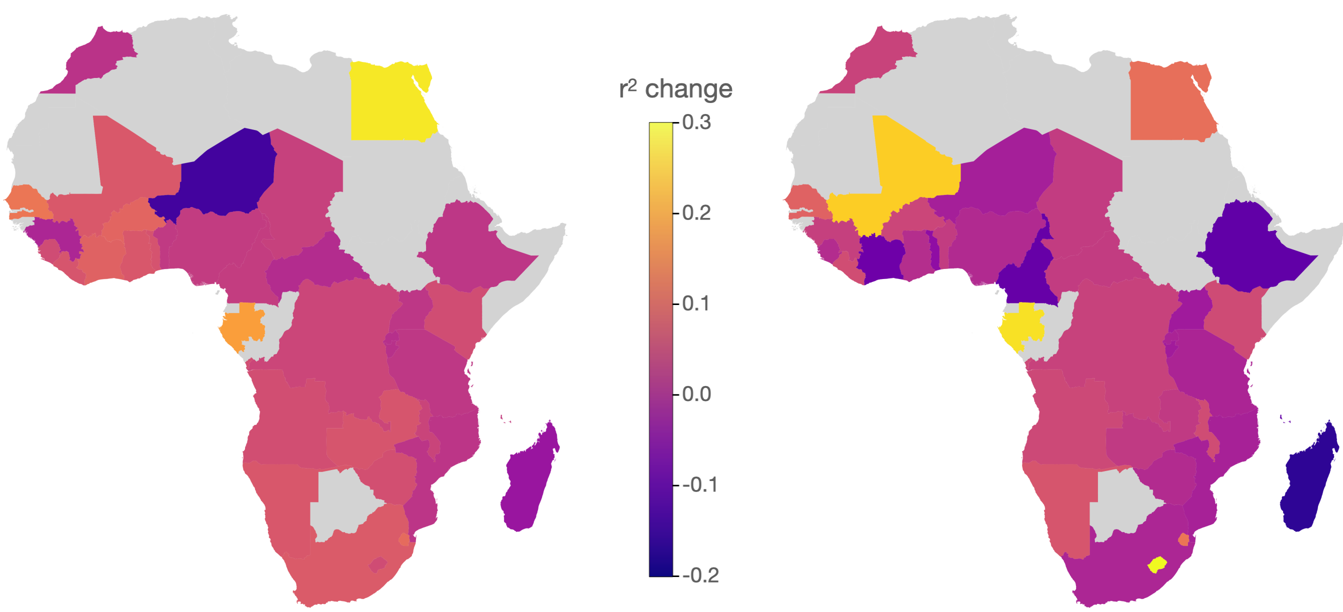
To verify the validity of our maps described in Section 5 and highlight the possibilities available when creating maps of this scale, we compared our predictions at a national level to the corresponding HDI. This was done by aggregating predictions from all patches within a country weighted by the population living there. This was done for each year from 2000 to 2019 using population rasters developed by Bondarenko *et al.* (2020). The resulting values represent a predicted wealth for the entire country which, when divided by the total population, can be considered something of a proxy for the mean IWI for all households. Each country year pairing was then compared to the corresponding Human Development Index (HDI). Even though HDI measures other forms of poverty than asset wealth, our aggregated country-year predictions are still able to explain 54% of the surveyed variation (Figure 8).



(a) Neighborhood level r^2

(b) District level r^2

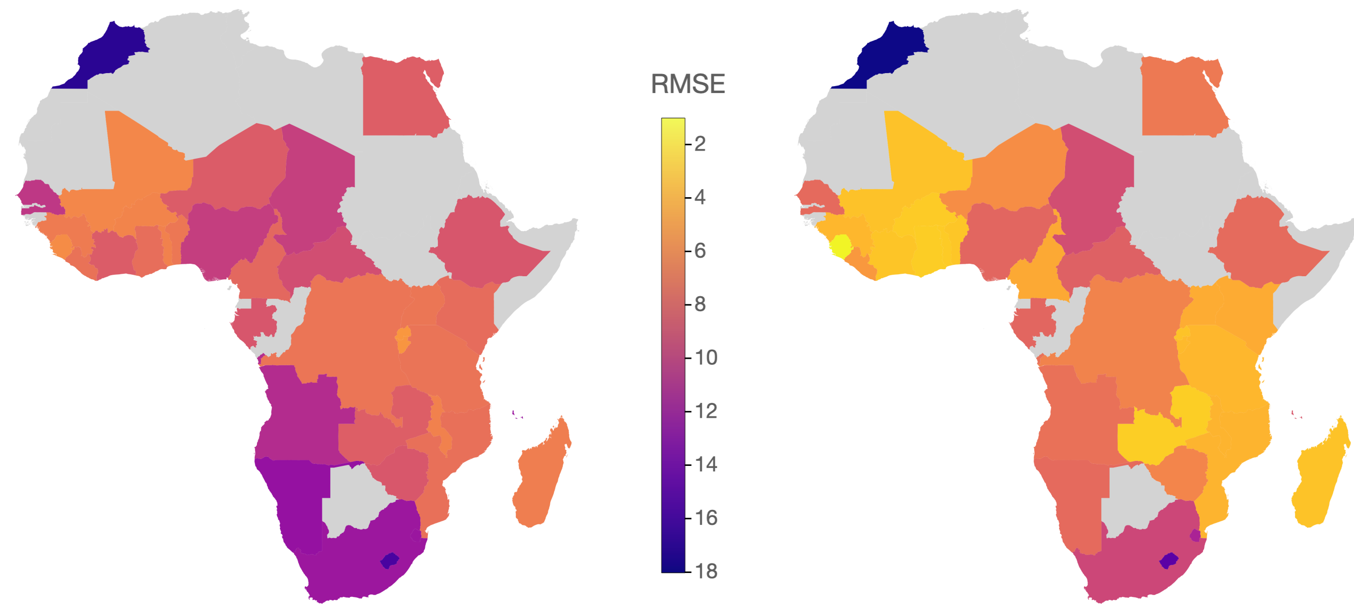
Figure 9: r^2 in held-out countries for the TEN-FRAME model. Grey countries do not have any survey data to compare with. (a) Predictions made at a cluster level has a pooled r^2 of 0.72. (b) Predictions made at a district level has a pooled r^2 of 0.78.



(a) Neighborhood level change to r^2

(b) District level change to r^2

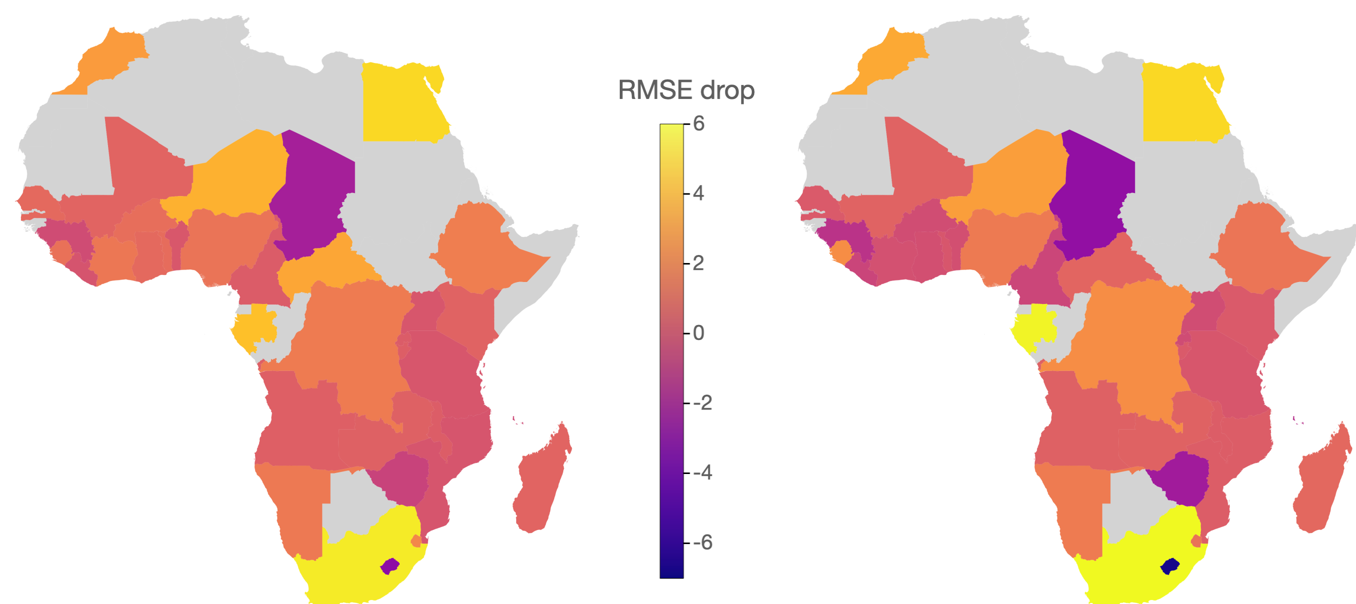
Figure 10: Changes to r^2 in held-out countries when switching from the SINGLE-FRAME model to the TEN-FRAME model. (a) The pooled r^2 change to predictions made at the cluster level is 0.12. (b) The pooled r^2 change to predictions made at the district level is 0.09.



(a) Neighborhood level RMSE

(b) District level RMSE

Figure 11: RMSE in held-out countries for the TEN-FRAME model. These values can be thought of as the mean number of IWI points the model is off when making predictions. (a) Predictions made at a cluster level has a RMSE of 8.04. (b) Predictions made at a district level has a RMSE of 5.83.



(a) Neighborhood level drop in RMSE

(b) District level drop in RMSE

Figure 12: Changes to RMSE in held-out countries when switching from the SINGLE-FRAME model to the TEN-FRAME model. (a) The drop in RMSE to predictions made at the cluster level is 1.35. (b) The drop in RMSE to predictions made at the district level is 1.2.

Fold	Countries	# clusters
A	Burundi, Egypt, Niger, South Africa, Zambia	11415
B	Benin, Gabon, Liberia, Mozambique, Nigeria, Uganda, Zimbabwe	11372
C	Burkina Faso, Cameroon, Chad, Comoros, Lesotho, Malawi, Senegal, Togo	11503
D	C.A.F., Côte d'Ivoire, D.R.C., Ethiopia, Guinea, Mali, Sierra Leone, Tanzania	11520
E	Angola, Eswatini, Ghana, Kenya, Madagascar, Morocco, Namibia, Rwanda	11385

Table 1: Table showing folds used for the out-of-country experiments

Fold	Time span	# clusters
A	1991-2003	11853
B	2004-2008	11571
C	2009-2012	10503
D	2013-2015	13266
E	2016-2019	10002

Table 2: Table showing folds used for the out-of-time-span experiments

Train	Val	Test
CDE	B	A
ADE	C	B
ABE	D	C
ABC	E	D
BCD	A	E

Table 3: Table showing information about the surveys used in this article.

Country	# clusters	IWI		# households per cluster	Survey years	Urban %
		mean	std			
Angola	969	32.68	24.06	23.2	06, 11, 15	50.2
Benin	1710	28.39	13.21	22.5	96, 01, 11, 17	44.2
Burkina Faso	1740	24.15	11.35	21.6	93, 98, 03, 10, 14, 17	31.1
Burundi	1128	19.03	10.84	26.1	10, 12, 16	19.2
Cameroon	1613	35.35	16.80	23.7	91, 04, 11, 18	52.9
C.A.F.	225	16.23	7.76	16.7	94	46.7
Chad	624	16.46	9.93	29.1	14	26.1
Comoros	242	38.37	12.38	17.7	12	43.8
Côte d'Ivoire	725	33.11	14.60	23.6	94, 98, 11	56.0
D.R.C.	783	18.43	14.13	32.2	07, 13	35.9
Egypt	7638	59.50	11.87	16.1	92, 95, 00, 03, 05, 08, 14	48.2
eSwatini	270	38.00	13.50	17.6	06	40.4
Ethiopia	2227	17.25	14.95	24.1	00, 05, 11, 16	29.0
Gabon	332	41.42	16.09	33.6	12	55.1
Ghana	2325	34.98	16.16	21.2	93, 98, 03, 08, 14, 16, 19	44.9
Guinea	1273	29.12	15.30	17.4	99, 05, 12, 18	36.1
Kenya	2616	26.67	14.18	24.3	03, 08, 14, 15	37.9
Lesotho	1162	26.77	12.26	21.2	04, 09, 14	27.1
Liberia	772	20.46	12.11	28.9	09, 11, 13, 16	42.4
Madagascar	1743	19.57	12.06	29.3	97, 08, 11, 13, 16	26.0
Malawi	3184	19.68	9.31	26.8	00, 04, 10, 12, 14, 15, 17	20.5
Mali	1941	26.47	13.16	22.1	95, 01, 06, 12, 15, 18	33.0
Morocco	476	54.65	17.09	21.7	03	56.7
Mozambique	1136	27.34	17.15	23.4	11, 15, 18	42.1
Namibia	1301	37.43	20.98	20.1	00, 06, 13	44.7
Niger	330	18.22	12.08	9.4	92, 98	57.6
Nigeria	4074	36.33	17.79	32.9	03, 08, 10, 13, 15, 18	39.3
Rwanda	1685	20.64	8.70	23.9	05, 07, 10, 11, 14	21.4
Senegal	2165	35.37	16.78	16.4	92, 97, 05, 08, 10, 12, 15, 19	41.0
Sierra Leone	1678	24.23	13.98	23.7	08, 13, 16, 19	36.3
South Africa	746	58.23	14.43	14.8	16	62.2
Tanzania	2668	23.24	13.57	17.3	99, 07, 10, 11, 15, 17	26.6
Togo	773	28.69	14.74	26.5	98, 13, 17	41.3
Uganda	1947	25.36	13.50	20.3	00, 06, 10, 11, 14, 16, 18	29.0
Zambia	1573	27.14	17.91	24.5	07, 13, 18	39.2
Zimbabwe	1401	32.83	17.82	23.3	99, 05, 10, 15	38.1
Total	57195	32.17	19.45	22.6	-	37.8

Table 4: Table showing information about the survey data used in this article.